

12

**(51) Int. Cl.<sup>5</sup>: H04N 5/232, G01S 3/786**

(74) Representative : Williams, Janice et al  
D. Young & Co., 21 New Fetter Lane  
London EC4A 1DA (GB)

[illegible]

This invention relates to a video camera system which can be applied in autonomous target tracking apparatus in which the field of view of a video camera can track the center of an object such as a human face.

Heretofore in video camera autonomous target tracking apparatus, a first autonomous target tracking method has been proposed in which the peak value of a high frequency element of a brightness signal in the measuring frame is memorized and its movement is automatically tracked as the feature of the target.

A second autonomous target tracking method has been proposed in which, regarding brightness signals of the front and back fields in the measuring frame, a motion vector is formed by taking the key point matching and it is assumed that the motion vector in the measuring frame is the movement of the target.

The first autonomous tracking method has an advantage in that the general construction is simple, but, for example in the case where there is a man in the measuring frame with a tree in the background, the man cannot be tracked automatically since the high frequency element of the tree is the biggest peak at this point.

Furthermore, since basically the first autonomous tracking method utilizes peak signals, this method is easily affected by noise. As a result, there is a possibility that automatic tracking cannot be performed in a photographic environment with low brightness. Also, it is possible that an object with poor contrast cannot automatically be tracked since basically the high frequency element is extracted.

Using the second autonomous tracking method, it is difficult to distinguish whether the calculated motion vector is caused by unintentional movement of the video camera or by the movement of the object, and in practice there is a possibility of the system functioning erroneously.

According to one aspect of the invention there is provided a video camera system comprising: an area dividing means (16, SP1) for dividing a display screen PIC to be shown by picture output signal S1 outputted from a lens block unit 1 into divided areas AR of the prescribed numbers 32 x 32, 16 x 16, a skin color divided area extracting means (16, SP1) for extracting skin color divided areas ARB in areas AR wherein hue signal HUE is skin color part, a lightness area extracting means (16, SP1) for extracting lightness divided area ARA in whole divided areas AR wherein brightness signal Y has the prescribed lightness, an absorbing means (16, SP2) for forming skin divided area groups FC1, FC2 and lightness divided area groups HR1, HR2 by absorbing blank areas around the skin divided area ARB and lightness divided area ARA, an adjacent divided area group extracting means (16, SP5) for extracting a set of adjacent skin divided area group FC1 and lightness divided area group HR1 in

skin divided area groups FC1, FC2 and lightness divided area groups HR1, HR2, and a measuring frame setting means (16, SP9) for setting said measuring frame PIC in order that a set of the skin divided area group FC1 and the lightness divided area group HR1 are included in the measuring frame PIC.

When blank areas AR around the skin divided area ARB and the lightness divided area ARA are absorbed, even though blank areas occur in the skin divided area ARB and the lightness divided area ARA affected by the intensity and/or direction of the light source, the skin divided area and the lightness divided area can be extracted properly because they are absorbed in skin divided area groups FC1, FC2 and lightness area groups HR1, HR2, and simultaneously, the measuring frame for the object can be set with certainty because the measuring frame is set based on the adjacent set FC1, HR1 in said skin divided area groups FC1, FC2 and the lightness divided area groups HR1, HR2.

Respective further aspects of the invention are as set forth in claims 1, 4, 6, 7, 8 and 9.

In embodiments of the invention, since the image information of each pixel which constitutes the display screen is divided into the prescribed number of small divided areas, and the measuring frame is set by judging the position and size of the skin color area based on the hue signal of each small divided area, a video camera system which is capable of setting the measuring frame on the object with high accuracy can be easily obtained.

Embodiments of the invention provide video camera autonomous tracking apparatus which is in general unaffected by the intensity and direction of the light source and, simultaneously, which is in general unaffected by the background.

Embodiments of the invention will now be described, by way of example, with reference to the accompanying drawings in which:

Fig. 1 is a block diagram showing the construction of video camera autonomous target tracking apparatus applicable to the video camera system;

Fig. 2 is a schematic diagram illustrating a HLS color coordinate system which expresses the visual stimulation;

Fig. 3 is a schematic diagram showing a display screen PIC wherein the image information can be obtained by dividing into small areas;

Fig. 4 is a flow chart showing the tracking process procedure;

Fig. 5 is a flow chart showing the tracking process procedure continued from Fig. 4;

Fig. 6 is a schematic diagram showing the display screen PIC having hair areas and skin color areas detected;

Fig. 7 is a schematic diagram showing the display screen PIC absorbed and/or combined; and

Fig. 8 is a block diagram showing a further embodiment of video camera target tracking apparatus.

In Fig. 1, ATT generally shows an autonomous target tracking apparatus of the video camera system, and the autonomous target tracker aims at a human face as its target. Thus, it is applicable as a means to make a lens block of the video camera system constituting the Autonomous Target Tracker with Human Recognition, ATT-HR system for automatically tracking function.

The autonomous target tracking apparatus ATT receives an image light LA which comes from the human face model as the object at a charge coupled device 4 (CCD) through a lens 2 of the lens block unit 1 and an iris 3 and delivers an image output signal S1 which shows the target image to a signal separation/automatic gain adjusting circuit 5.

The signal separation/automatic gain adjusting circuit sample holds the image output signal S1, which is gain controlled in order to have the prescribed gain by the control signal from an auto iris (AE) system (not shown) and thus obtained image output signal S2 will be supplied to a digital camera processing circuit 7 through an analog-to-digital converter 6.

The digital camera processing circuit 7 forms brightness signal Y, chrominance signal C and color difference signals R-Y and B-Y according to the image output signal S2, and outputs the brightness signal Y and chrominance signal C through a digital-to-analog converter 8 as video signal S3.

Furthermore, the digital camera processing circuit 7 supplies brightness signal Y and color difference signals R-Y and B-Y as target tracking detection signal S4 to a tracking control circuit 11, which generates tracking control signal S5 for a panning driving motor 12 and a tilting driving motor 13 which are equipped in the lens block unit 1 in accordance with the target tracking detection signal S4.

First of all, the tracking control circuit unit 11 supplies color difference signals R-Y and B-Y to a saturation/hue detection circuit 14 and forms the hue signal HUE and the saturation signal SAT and memorizes these with brightness signal Y in an image memory 15 constituted by, such as field memory, per each picture element unit.

The saturation/hue detection circuit 14 forms hue signal HUE and saturation signal SAT by rectangular coordinate to polar coordinate converting color difference signals R-Y and B-Y, and recognizes the human face model as an object based on the visual stimulation which human being can perceive by means of brightness signal Y, hue signal HUE and saturation signal SAT.

In this connection, as shown in Fig. 2, the visual stimulation which human beings can generally perceive, can be expressed in L axis and color coordinate system which contains SH plane perpendicularly in-

tersecting to L axis, as it is called "HLS system". L axis shows "lightness" and this is equivalent to brightness signal Y. SH plane is expressed by polar coordinate which intersects perpendicularly to L axis. In the SH plane, S shows "saturation" and is expressed by the distance from L axis. Also, H stands for "hue" and hue is expressed by an angle at the time when the direction of hue signal B-Y is 0 [°].

The cubic factor of this HLS is represented that, as the light source becomes lighter, the color coordinate, i.e., SH plane, moves upward along the L axis, and all colors become white. At this point, the saturation S gradually decreases. On the other hand, if the light source becomes darker, the color coordinate, i.e., SH plane, falls down along the L axis, and all colors become black. At this point, the saturation S also decreases.

As described above, the saturation S and the brightness Y are easily affected by the lightness of the light source in accordance with the characteristic of HLS color coordinate system; and therefore, it is difficult to say that they are the optimal parameters to express the quantity of features of the object. On the other hand, it is obvious that the hue H is the quantity which is not easily affected by the light source as a parameter to express the quantity of inherent features of the object.

However, in the case where color of the object is in the neighborhood of L axis, i.e., in the case of white, black or gray, the signal of hue H has no meanings as the information, and in the worst case, in spite of the fact that the picture with poor S/N is white, it seems possible to have various hue H vectors.

In utilizing these features of HLS color coordinate system, the tracking control circuit 11 extracts the feature of the human face model as the object and when the feature changes, the panning driving motor 12 and tilting driving motor 13 will be actuated in order to follow the changes and as a result, video signal which tracks the movement of the object as video signal S3 will be obtained.

More specifically, target tracking detection signal S4 memorized in an image memory 16 supplies address signal S7 to divide the display screen PIC formed in the image memory 15 into divided areas AR of the prescribed size in accordance with xy rectangular coordinates (x, y), as shown in Fig. 3, by giving block defining signal S6 to be outputted from the microprocessor composed tracking signal processing circuit 16 to an address generation circuit 17.

Thus, the data of each picture element which constitute the display screen PIC of the image memory 16 will be read out in every divided area AR and will be evaluated as an image information in each divided area AR.

In the case of this embodiment, the display screen PIC is divided into 32 (or 16) small divided areas AR in the directions of x and y respectively, and

thus by designating the coordinates of rectangular coordinate (x, y) as  $x = i$ ,  $y = j$ , regarding  $32 \times 32$  (or  $16 \times 16$ ) small divided areas AR, the image information I ( $x = i$ ,  $y = j$ ) of small divided areas to be designated can be read out.

With this arrangement, whereas the hue signal HUE element will be given to a skin color detection circuit 19 through a gate circuit 18 in the image information I ( $x = i$ ,  $y = j$ ) to be read out in every small divided area from the image memory 15, the brightness signal Y element will be given directly to a hair color detection circuit 20.

The skin color detection circuit 19 detects the image part of skin of the human face model and when the incoming hue signal HUE element is within the range of prescribed skin color, it gives "+1" count output S8 to a skin color pixel number counter 21 as the image of said picture element is skin color.

In practice, the dynamic range of hue signal HUE takes the value in the range of an angle  $0 - 350 [^\circ]$ .

On the other hand, the hair color detection circuit 20 detects hair areas in the image part of the human face model and in the case of this embodiment, the dynamic range of brightness signal Y is expressed by the value  $0 - 255$  (8 bits) and when the brightness signal of each picture element is below the value 50, it defines as black and judges that said picture element exists in the hair area and outputs "+1" count output S9 to a hair pixel number counter 22.

With this arrangement, the number of pixels having information on skin color and hair color respectively in pixels contained in each divided area AR, will be counted in the skin color pixel number counter 21 and the hair pixel number counter 22.

The counting result of each divided area AR of the skin color pixel number counter 21 and the hair pixel number counter 22 will be compared with judgement reference signals S10 and S11 to be given from the tracking signal processing circuit 16 at comparators 23 and 24.

At this point, judgement reference signals S10 and S11 are set to the value to determine whether the counting result of skin color pixel number counter 21 and hair pixel number counter 22 will be judged as prospective skin color area or prospective hair area or not for said divided areas AR, and thus, comparators 23 and 24 transmit prospective skin color area detection information S12 and prospective hair area detection information S13 to the tracking signal processing circuit 16.

In the case of this embodiment, a hue noise gate signal forming circuit 25 composed of comparator is equipped for the gate circuit 18 and it compares saturation signal SA1 to be read out in every pixel from the image memory 15 with noise judgement signal S14 to be outputted from tracking signal processing circuit 16, and by giving gate signal S15 which operates the gate circuit to close when saturation signal

SAT is under the prescribed level, it is so arranged that the hue signal HUE element of said pixel will not be inputted to the skin color detection circuit 19.

In this connection, in the case where the hue signal HUE detected at the saturation/hue detection circuit 14 is in the neighborhood of the L axis (Fig. 2) and this means that said hue signal HUE has small saturation and for fear that it is buried in noises and has no meanings as information, these meaningless hue signals will be eliminated at the gate circuit 18.

According to the foregoing construction, the tracking signal processing circuit 16 judges the feature of an image on the display screen PIC from the prospective skin color area detection information S12 and the prospective hair area detection information S13 to be obtained based on brightness signal Y, hue signal HUE and saturation signal SAT of each pixel contained in the image memory 15 and delivers tracking control signal S5 so as the center of gravity of human face model comes in the center of the measuring frame constantly, to the panning driving motor 12 and the tilting driving motor 13.

In the case of this embodiment, supposing that the display screen PIC is generally one measuring frame, it is so arranged that the centroid of a face model would come in the center of display screen PIC.

When the tracking signal processing circuit 16 enters in the tracking processing procedure RT1 of Fig. 4, first of all, it sets the field number FN to the initial value,  $FN = 1$ , and moves to step SP1 and divides the display screen PIC into  $32 \times 32$  divided areas AR, and then obtains counting results of the skin color pixel number counter 21 and the hair pixel number counter 22 based on said detection result of the skin color detection circuit 19 and the hair detection circuit 20 for pixels contained in each divided area AR. In utilizing these the tracking signal processing circuit 16 takes into the prospective skin color area detection information S12 when said divided area is the prospective skin color area and at the same time it takes into the prospective hair area detection information S13 when the divided area AR is the prospective hair area.

When the above procedure has been terminated at the step SP1, the tracking signal processing circuit 16 appoints the field number to be extracted and processed next time by incrementing the field number  $FN$  to  $FN = FN + 1$ .

With this arrangement the procedure of step SP1 has been terminated, and the tracking signal processing circuit 16 is able to obtain the distribution condition of divided areas AR judged as the prospective hair divided areas ARA and prospective skin color divided areas ARB as shown in Fig. 3 or Fig. 6.

At this point, as shown in Fig. 3, if the prospective hair divided areas ARA and the prospective skin color divided areas ARB are adjacent each other without space and moreover, the extraction result showing

that there is no space between the prospective hair divided areas ARA and the prospective skin color divided areas ARB is obtained, it means that the information to conform to the actual condition of human face model as an object has been obtained.

However, since in ordinary circumstances the direction and reflection manner of the light source for the object are not uniform, there occur blank areas in the prospective hair divided areas ARA and prospective skin color divided areas ARB as shown in Fig. 6, and moreover, there are many cases where extracting results show that there are blank areas in the group of prospective hair divided areas ARA and the group of prospective skin color divided areas ARB.

Then, the tracking signal processing circuit 16 absorbs said discontinuous areas according to (hair area absorbing rule 1) and (skin color area absorbing rule 2) at the following step SP2.

The "absorption rule 1" of discontinuous areas is "If  $I(i + p, j + q) = \text{hair area}$ , ( $p, q = +1$  or  $-1$  or  $0$ , provided excepting  $p = q = 0$ ), then absorb  $I(i, j)$  as the prospective hair area".

This "absorption rule 1" of discontinuous areas means that in the address  $(x = i, y = j)$  area the image information  $I(i, j)$  of said designated divided areas will be absorbed in the prospective hair divided areas ARA in the case where there are prospective hair divided areas ARA in the neighboring areas centering around this divided areas AR.

For example, in Fig. 6 in the case where the blank divided area  $(x = 6, y = 5)$  is designated as an appointed area, the hair divided area  $(x = 6, y = 4)$  is included as prospective hair divided area ARA1 in divided areas AR, i.e.,  $(x = 5, y = 4)$ ,  $(x = 6, y = 4)$ ,  $(x = 7, y = 4)$ ,  $(x = 7, y = 5)$ ,  $(x = 7, y = 6)$ ,  $(x = 6, y = 6)$ ,  $(x = 5, y = 6)$  and  $(x = 5, y = 6)$  which surround this blank divided area, and thus this blank divided area  $(x = 6, y = 5)$  is absorbed in a group of prospective hair divided areas ARA1 by the "absorption rule 1".

Also, the absorption rule of discontinuous area will be applied after applying the "absorption rule 2", and its content is as follows: If  $I(i + p, j + q) = \text{prospective skin color divided area}$ ; ( $p, q = +1$  or  $-1$  or  $0$ , provided, excepting  $p = q = 0$ ), then  $I(i, j)$  is absorbed as prospective skin color divided area.

The "absorption rule 2" of this discontinuous area is the same as described above regarding prospective hair divided area ARA. If there are prospective skin color areas in the neighbor areas concerning the designated divided area  $(x = i, y = j)$ , the image information  $i(x = i, y = j)$  of said designated divided area  $(x = i, y = j)$  will be absorbed as prospective skin color divided area ARB.

The "absorption rule 2" of this discontinuous area is the same as described above regarding prospective hair divided area ARA. If there are prospective skin color areas in the neighbor areas concerning the designated divided area  $(x = i, y = j)$ , the image infor-

mation  $i(x = i, y = j)$  of this designated divided area  $(x = i, y = j)$  will be absorbed as prospective skin color divided area ARB.

For example, in Fig. 6 in the case where the blank divided area  $(x = 6, y = 6)$  is appointed as a designated area, since the skin color area  $(x = 7, y = 7)$  is included as prospective skin color divided area ARB1 in divided areas AR surrounding this blank divided area  $(x = 6, y = 6)$ , i.e.,  $(x = 5, y = 5)$ ,  $(x = 6, y = 5)$ ,  $(x = 7, y = 5)$ ,  $(x = 7, y = 6)$ ,  $(x = 7, y = 7)$ ,  $(x = 6, y = 7)$ ,  $(x = 5, y = 7)$  and  $(x = 5, y = 6)$ , this blank divided area  $(x = 6, y = 6)$  is absorbed in a group of prospective skin color areas ARB1 by the "absorption rule 2".

Accordingly, the absorption process of discontinuous area of the step SP2 is terminated and as shown in Fig. 6, since there were blank areas in the group of prospective hair divided areas ARA1, the group of prospective hair divided areas ARA1 was discontinuous, however, this blank areas can be filled in by the absorbed prospective skin color divided area ARB2 as shown in Fig. 7.

The tracking signal processing circuit 16 proceeds to the step SP3 when the process of step SP2 is terminated and defines the determination of hair area and skin color area by the "hair area determination rule 1" and the "skin color area determination rule 2".

The "hair area determination rule 1" is "If  $I(i, j) = \text{prospective hair area}$  and  $I(i + p, j + q) = \text{prospective hair area}$ , (here  $p, q = +1$  or  $-1$  or  $0$ , provided, excepting  $p = q = 0$ ), then  $I(i, j), I(i + p, j + q)$  will be combined as hair area.

This "hair area determination rule 1" means that in the address  $(x = i, y = j)$  area, the image information  $(i, j)$  of said designated area  $(x = i, y = j)$  will be combined in the hair divided area group HR, in the case where there are prospective hair divided areas in the neighboring divided areas centering around said divided areas.

Also, the "skin color area determination rule 2" combines  $I(i, j), I(i + p, j + q)$  as skin color area, "If  $I(i, j)$  is prospective skin color area and  $I(i + p, j + q)$  is prospective skin color area, (here  $p, q = +1$  or  $-1$  or  $0$ , provided, excepting  $p = q = 0$ )".

This "skin color area determination rule 2" is the same as described above regarding the prospective hair divided area, the image information  $(x = i, y = j)$  and the image information of neighboring area where there exists the prospective skin color divided area group FC if there are prospective skin color areas in the neighboring areas regarding the designated divided area  $(x = i, y = j)$ .

Then, for example, as described above, if the "hair area determination rule 1" is applied to the area  $(x = 6, y = 5)$  absorbed in the hair divided area (Fig. 6) by the "hair area absorption rule 1", hair divided areas  $(x = 5, y = 5)$ ,  $(x = 5, y = 4)$ ,  $(x = 6, y = 4)$ ,  $(x = 7, y = 4)$  and  $(x = 7, y = 5)$  are included as prospective

hair divided areas in divided areas around the divided area  $(x = 6, y = 5)$ , i.e.,  $(x = 5, y = 4)$ ,  $(x = 6, y = 4)$ ,  $(x = 7, y = 4)$ ,  $(x = 7, y = 5)$ ,  $(x = 7, y = 6)$ ,  $(x = 6, y = 6)$ ,  $(x = 5, y = 6)$  and  $(x = 5, y = 5)$ , and areas  $(x = 5, y = 4)$ ,  $(x = 6, y = 4)$ ,  $(x = 7, y = 4)$ ,  $(x = 7, y = 5)$ ,  $(x = 6, y = 5)$  and  $(x = 5, y = 5)$  will be combined as the hair divided area group FC1.

Moreover, for example, as described above, if the "skin color area determination rule 2" is applied to the divided area  $(x = 6, y = 6)$  absorbed in the skin color divided area by the "skin color area absorption rule 2", areas  $(x = 6, y = 6)$ ,  $(x = 7, y = 6)$ ,  $(x = 7, y = 7)$  and  $(x = 6, y = 7)$  will be combined as the skin color divided area group FC1 because skin color divided areas  $(x = 7, y = 6)$ ,  $(x = 7, y = 7)$  and  $(x = 6, y = 7)$  are included as prospective skin color areas in divided areas around the divided area  $(x = 6, y = 6)$ , i.e.,  $(x = 5, y = 5)$ ,  $(x = 6, y = 5)$ ,  $(x = 7, y = 5)$ ,  $(x = 7, y = 6)$ ,  $(x = 7, y = 7)$ ,  $(x = 6, y = 7)$ ,  $(x = 5, y = 7)$  and  $(x = 5, y = 6)$ .

Thus, as shown in Fig. 7, divided area groups HR1 and HR2 in the display screen PIC are combined as the hair area and divided area groups FC1 and FC2 are combined as the skin color area at the steps SP2 and SP3 by absorbing discontinuous divided areas and by combining hair divided area and skin color divided area.

Then, the tracking signal processing circuit 16 proceeds to the step SP4 of Fig. 5 and calculates the sizes Area\_Hair and Area\_Skin of combined hair areas HR1 and HR2, and skin color areas FC1 and FC2.

In the case of display screen PIC, as shown in Fig. 7, the size Area\_Hair of the hair area HR1 is calculated as the total number of areas contained in the hair area HR1, i.e.  $\text{Area\_Hair} = 45$ . In the same manner, the size Area\_Hair of the hair area group HR2 is calculated as the  $\text{Area\_Hair} = 14$ , the size Area\_Skin of skin color area group FC1 is calculated as  $\text{Area\_Skin} = 66$  and the size Area\_Skin of skin color area group FC2 is calculated as  $\text{Area\_Skin} = 10$ .

Then, the tracking signal processing circuit 16 proceeds to the step SP5 and judges the process to determine the face area under the condition to satisfy conditions of the "face area determination rule 1" and the "face area determination rule 2".

The "face area determination rule 1" identifies one hair divided area group and one skin color divided area group as a set of areas and if the size ratio of the size Area\_Hair of the one hair area group and the size Area\_Skin of one skin color area group, i.e.,  $\text{Area\_Hair}/\text{Area\_Skin} > 5$  and  $\text{Area\_Hair}/\text{Area\_Skin} < 1/5$ , determines as prospective face area.

This "face determination rule 1" judges that this set of divided area group is most possible the face area if the size ratio of the sizes Area\_Hair and Area\_Skin is less than five times and more than 1/5 times.

The above judgement avoids the possibility of

misjudging that all dark areas would be judged as hair area, such as the image having numerous divided areas with dark brightness level on the display screen PIC (e.g., in the case where there is a dark screen in the back), similarly, in the case where there exist a large quantity of skin color divided areas besides face skin, it is possible that the judgement of face area can not correctly be defined, and the above judgement of face area may not correctly be defined, and the above judgement thus avoids erroneous judgement.

Furthermore, the "face area determination rule 2" defines that "if  $I(i, j) = \text{hair area}$  and  $I(i, j-1) = \text{skin color area}$ , Neighbor = Neighbor + 1. If Neighbor > 3, then hair area + skin color area is face".

This "face area determination rule 2" judges a set of hair area group and skin color area group as the face if the upper side of the screen is hair divided area group (i.e.,  $I(i, j)$ ) and the lower side is skin color area group (i.e.,  $I(i, j-1)$ ) and under the condition that these hair divided area group and skin color divided area group are in contact with at least three divided areas.

According to this "face area determination rule 2", a pair of divided area groups, i.e., a set of hair area group HR1 and skin color area group FC1 is in contact with 8 divided areas and satisfies this rule 2, and therefore, it is judged as a face.

On the other hand, because a pair of divided area groups, i.e., a set of hair area group HR2 and skin color area group FC2 is not in contact with any areas and does not satisfy the condition of rule 2, it is excluded from the judgement of face area.

Furthermore, the tracking signal processing circuit 16 proceeds to the step SP6 and judges whether the process of steps SP2 - SP5 has been repeated twice or not. When the first process is finished, it proceeds to the step SP7 and changes the divided number divided at the step SP1 from  $32 \times 32$  to  $16 \times 16$  and extracts prospective hair area and prospective skin color area and returns to the step SP2.

Thus, the tracking signal processing circuit 16 executes steps SP2 - SP3 - SP4 - SP5 on the divided areas of  $16 \times 16$ , and thus improve the probability of face recognition by executing the process on roughly divided areas than the process on  $32 \times 32$  divided.

In this connection, there are cases where the face object which could not be recognized at the time of  $32 \times 32$  divided process can be recognized by processing on the roughly divided areas. For instance, it is especially profitable in case of treating a zoom uped large face in the display screen PIC.

When an affirmative result is obtained at the step SP6, it means that the image processing has been finished twice and the tracking signal processing circuit 16 proceeds to the step SP8 and when more than two faces are extracted from one display screen PIC, it executes the process so as to select the face area near the center of display screen PIC as an object at the time when the record pause button is released,

and then at the step SP9, it supplies tracking control signal S5 in order that the centroid of face will be the center of the measuring frame, i.e., the display screen PIC, to the panning driving motor 12 and the tilting driving motor 13.

Thus, the tracking signal processing circuit 16 terminates a cycle of the automatic tracking process and returns to the step SP1 mentioned above to enter the following automatic tracking cycle. With this arrangement, since it is used as quantity of a characteristic a hue signal which is hard to be affected by an intensity and a direction of light source and the most effectively distinguishes the characteristic of the object on the display screen, an autonomous target tracking apparatus which is hard to be affected by the background and having resistance to the movement of the hands can be obtained.

In this connection, in the case where the characteristic quantity of the object is processed based on the brightness signal as before, if the lightness of the display screen PIC brightens up extremely or becomes dimmer, as described above regarding the HLS color coordinate system in Fig. 2, it has been difficult to grasp the image information of the object. However, in the case of the above embodiment, even through the lightness becomes extremely brighter or dimmer, the characteristic quantity of the object cannot be lost.

Also, according to the foregoing construction, since the peak value has not been used as the characteristic quantity, it is resistant to noises and since the human face model which is not affected by changes of the coordinate on the screen is used without using the motion vector, the system which has resistance to the movement with hand of camera will be constructed.

Furthermore, according to the foregoing construction, since it holds the human face model inside, the user does not have to initialize the object and there is no need for setting up the measuring frame and if there is a background in the measuring frame, the phenomenon that a center of a picture would be pulled by this cannot be occurred.

Fig. 8 illustrates the other embodiment according to the present invention and as shown, the identical reference numerals are given with corresponding parts of Fig. 1, brightness signal Y, color difference signal R-Y and B-Y obtained from the digital camera processing circuit 7 will be given to the screen reduction circuit 32 via a low pass filter 31. The screen reduction circuit 32 executes the process to divide the display screen PIC into divided areas 32 x 32 or 16 x 16 by means of screen reduction control signal S21 outputted from the tracking signal processing circuit 16 and brightness signal Y and color difference signal R-Y and B-Y of each pixel outputted from the digital camera processing circuit 7.

Accordingly, hue signal HUE and saturation sig-

nal SAT will be supplied to the image memory 15 by giving color difference signals R-Y and B-Y corresponding to 32 x 32 areas or 16 x 16 areas from the screen reduction circuit 32 to the saturation/hue detection circuit 14 as target tracking detection signal S4 and simultaneously, brightness signal Y will be supplied directly to the image memory 15 as target tracking detection signal.

In this case, the image memory 15 has a memory area for 32 x 32 images, therefore, in the case of Fig. 1 it is necessary to have memory capacity for one field of pixel, on the other hand in the case of Fig. 8, it can be simplified further.

With this arrangement, hue signal HUE and saturation signal SAT will be outputted as signals accumulated in 32 x 32 pixels or 16 x 16 pixels, and thus the hair detection circuit 20 and skin color detection circuit 19 detect divided areas 32 x 32 or 16 x 16 whether they are the hair areas or skin color areas respectively and supply the detection result to the tracking signal processing circuit 16 as hair area detection output S22 and skin color area detection output S23.

According to the construction of Fig. 8, since it is so arranged that the saturation/hue detection circuit 14, the image memory 15, the hair detection circuit 20 and the skin color detection circuit 19 will be processed after the image signal for one screen to be outputted from the digital camera processing circuit 7 is reduced as image information of 32 x 32 divided areas or 16 x 16 divided areas at the screen reduction circuit 19, the circuit construction of said processing circuits can be further simplified.

In the case where the autonomous target tracking apparatus ATT composed of Fig. 1 or Fig. 8 is applied to an auto focus adjusting system, since the auto focus adjusting system tracks in order that the lens block unit 1 can set the measuring frame followed by the display screen PIC corresponding to the change of the face centroid, an automatic focus adjusting system which can focus the lens 2 on the centroid position of a face can be obtained.

Also, in the case of applying the autonomous target tracking apparatus of the above mentioned construction to an automatic iris adjusting system, since it can be so arranged that the lens block unit 1 can set the measuring frame on the centroid of the face followed by the display screen PIC, and the centroid of the face can be always light measured, it becomes unnecessary to correct the counter light as the auto iris adjusting system and simultaneously, the brightness level of the object can be set automatically to the optimum value.

Furthermore, in the case of applying the autonomous target tracking apparatus to an automatic white balance adjusting system, since the lens block unit 1 sets the measuring frame to track the centroid of the face followed by the display screen PIC, the auto white balance adjusting system can adjust colors



of overall screen by adjusting to get optimal skin color or the face. And as a result, the auto white balance adjusting system as a whole can adjust the auto skin color balance (ASB) and thus, as compared with the previously proposed auto white balance adjusting apparatus which adjusts because the integrated result of the entire screen is white, even in the case where the environment of the object is over exposed or under exposed, the display screen having excellent color balance can be obtained.

Furthermore, the embodiment described above has dealt with the case of executing absorption, combination and determination processes on the group of hair divided areas and the group of skin color divided areas by setting the measuring frame on the overall display screen PIC. However, the same effects as those of the embodiment described above can be obtained if the measuring frame is set in the area selected according to demands in the display screen PIC.

Moreover, the embodiment described above has dealt with the case of shifting the field of vision of the lens block 1 corresponding to the movement of the object by panning and/or tilting and thus the desired object image will be entered in the measuring frame. However, the same effects as those of the embodiment described above can be obtained by shifting the measuring frame for the image in the display screen PIC without panning or tilting.

It will be apparent to those skilled in the art that various changes and modifications may be made to the specific embodiments described above without departing from the scope of the invention.

## Claims

### 1. A video camera system comprising:

area dividing means (16) for dividing a display image to be shown by image output signals output from a lens block unit (1) into divided areas (AR) of prescribed numbers;

skin color area extracting means (16) for extracting skin color divided areas (ARB) wherein hue signals are skin color part in all of said divided areas;

lightness area extracting means (16) for extracting lightness divided areas (ARA) wherein brightness signal are in prescribed luminance in all of said divided areas;

absorbing means (16) for forming skin color divided area groups and lightness divided area groups by absorbing blank divided areas around said skin color divided area and said lightness divided area;

adjacent divided area extracting means (16) for extracting a set of skin color divided area group and lightness divided area group which are adjacent each other from said skin color divided

area groups and lightness divided area groups; and

measuring frame setting means (16) for setting said measuring frame in order that said set of skin color divided area group and lightness divided area group can be entered into said measuring frame.

2. A video camera system as claimed in claim 1, wherein said measuring frame setting means sets the image part of a set of said skin color divided area group and lightness divided area group within said measuring frame by moving said measuring frame on said display screen depending upon the extracted result of said adjacent divided area group extracting means.

3. A video camera system as claimed in claim 1, wherein said measuring frame setting means sets the image part of a set of said skin color divided area group said lightness divided area group within said measuring frame by panning and/or tilting said lens block unit depending upon the extracted result of said adjacent divided area group extracting means.

### 4. A video camera system comprising:

area dividing means (16) for dividing a display image to be shown by image output signals output from the lens block unit (1) into divided areas (AR) of the prescribed numbers;

skin color area extracting means (16) for extracting skin color divided areas (ARB) wherein hue signal is skin color part in all of said divided areas;

hair area extracting means for extracting hair divided areas wherein brightness signal has hair lightness in all of said divided areas;

absorbing means (16) for forming skin color divided area groups and hair divided area groups by absorbing blank divided areas around said skin color divided area and said hair divided area;

adjacent divided area extracting means (16) for extracting a set of skin color divided area group and hair divided area group which are adjacent each other from said skin color divided area groups and hair divided area groups;

face defining means for judging said set of skin color divided area group and hair divided area group to be a face under the condition that said hair divided area group is adjacent to the upper part of said skin color divided area group in said set of skin color divided area group and hair divided area group extracted in said adjacent divided area extracting means; and

measuring frame setting means (16) for setting said measuring frame in order that said



face area is entered into said measuring frame.

5. A video camera system as claimed in claim 4, wherein said measuring frame setting means (16) for setting said face area to enter in said measuring frame by panning and/or tilting said lens block unit (1) in order that the centroid of said face comes in the center of said display screen.

6. A video camera system comprising:  
area dividing means (16) for dividing a display image to be shown by image output signals output from a lens block unit (1) into divided areas (AR) of the prescribed numbers;

skin color area extracting means (16) for extracting skin color divided areas (ARB) wherein hue signal is skin color part in all of said divided areas;

hair area extracting means for extracting hair divided areas wherein brightness signal has lightness of hair in all of said divided areas;

absorbing means (16) for forming skin color divided area groups and hair divided area groups by absorbing blank divided areas around said skin color divided area and said hair divided area;

adjacent divided area group extracting means (16) for extracting a set of skin color divided area group and hair divided area group which are adjacent each other from said skin color divided area groups and hair divided area groups;

face defining means for judging said set of hair divided area group and skin color divided area group to be a face under the condition that the sizes of said hair divided area group and said skin color divided area group are calculated regarding said set of skin color divided area group and hair divided area group extracted in said adjacent divided area group extracting means, the ratio of said sizes of said hair divided area group and said skin color divided area group is within the limit of the prescribed numbers, and said hair divided area group is adjacent the upper side of said skin color divided area group; and

measuring frame setting means (16) for setting said measuring frame in order that said face area can be entered into said measuring frame.

7. An autonomous target tracking method for automatically tracking an object displayed on a display screen based on imaging output signals output from a lens block unit (1), said autonomous target tracking method comprising the steps of:

dividing a display image to be shown by said image output signals into divided areas (AR) of prescribed numbers;

extracting skin color divided areas (ARB) wherein hue signal is skin color part in all of said divided areas;

extracting lightness divided areas wherein brightness signal are in prescribed luminance in all of said divided areas;

absorbing blank divided areas around said skin color divided area and said lightness divided area to form skin color divided area groups and lightness divided area groups;

extracting a set of skin color divided area group and lightness divided area group which are adjacent each other from said skin color divided area groups and lightness divided area groups; and

setting said measuring frame in order that said set of skin color divided area group and lightness divided area group can be entered into said measuring frame.

8. An autonomous target tracking method for automatically tracking an object displayed on a display screen based on an imaging output signals output from a lens block unit (1), said autonomous target tracking method comprising the steps of:

dividing a display image to be shown by said image output signals into divided areas (AR) of prescribed numbers;

extracting skin color divided areas (ARB) wherein hue signal is skin color part in all of said divided areas;

extracting hair divided areas wherein brightness signal has hair lightness in all of said divided areas;

absorbing blank divided areas around said skin color divided areas and said hair divided areas to form skin color divided area groups and hair divided area groups;

extracting a set of skin color divided area group and hair divided area group which are adjacent each other from said skin color divided area groups and hair divided area groups;

judging said set of skin color divided area group and hair divided area group to be a face under the condition that said hair divided area group is adjacent to the upper part of said skin color divided area group in said set of skin color divided area group and hair divided area group; and

setting said measuring frame in order that said face area is entered into said measuring frame.

9. An autonomous target tracking method for automatically tracking an object displayed on a display screen based on imaging output signals output from a lens block unit, said autonomous target tracking method comprising the steps of:

dividing a display image to be shown by

said image output signals into divided areas (AR)  
of prescribed numbers;

extracting skin color divided areas (ARB)  
wherein hue signal is skin color part in all of said  
divided areas;

5

extracting hair divided areas wherein  
brightness signal has lightness of hair in all of said  
divided areas;

absorbing blank divided areas around said  
skin color divided areas and said hair divided  
areas to form skin color divided area groups and  
hair divided area groups;

10

extracting a set of skin color divided area  
group and hair divided area group which are ad-  
jacent each other from said skin color divided  
area groups and hair divided area groups;

15

judging said set of hair divided area group  
and skin color divided area group to be a face un-  
der the condition that sizes of said hair divided  
area group and said skin color divided area group  
are calculated regarding said set of skin color di-  
vided area group and hair divided area group, the  
ratio of said sizes of said hair divided area group  
and said skin color divided area group is within  
the limit of the prescribed numbers, and said hair  
divided area group is adjacent upper side of said  
skin color divided area group; and

20

25

setting said measuring frame in order that  
said face area can be entered into said measuring  
frame.

30

35

40

45

50

55

10

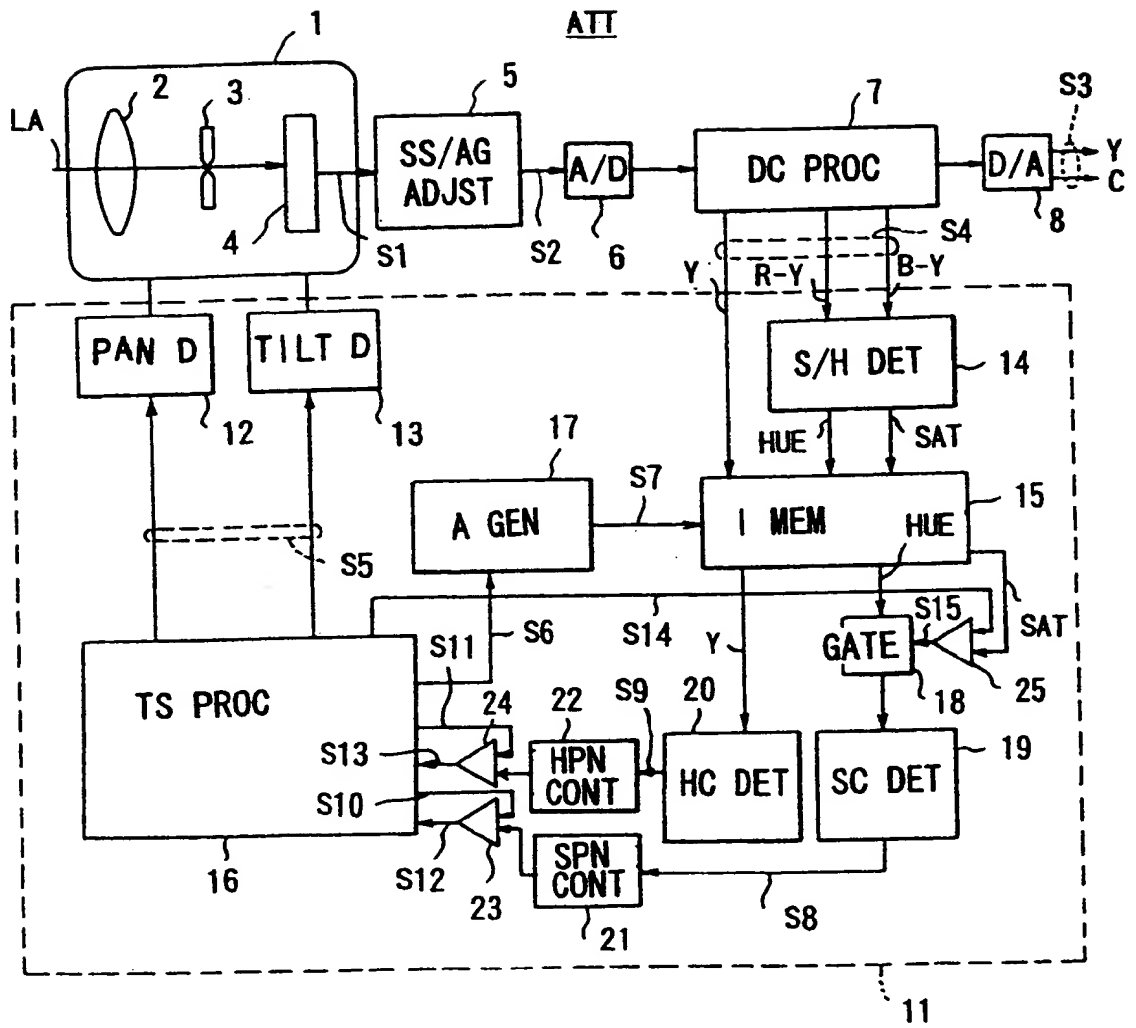


FIG. 1

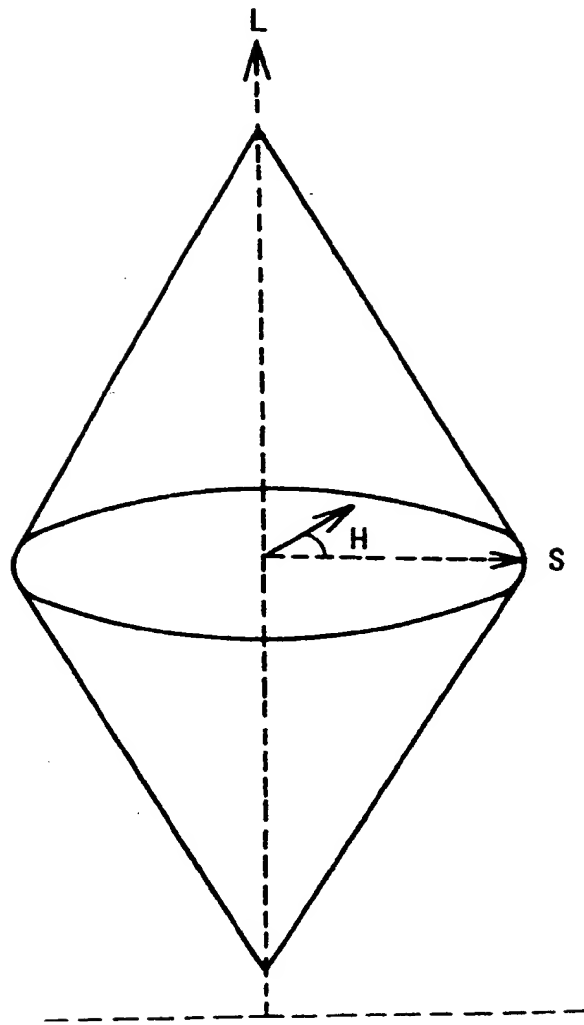


FIG. 2

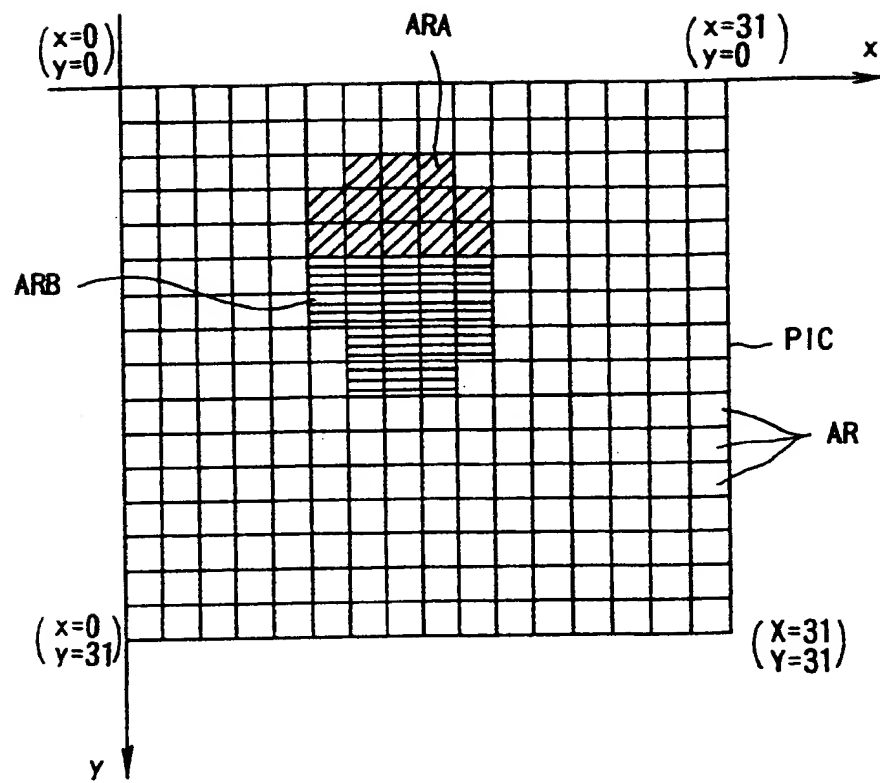


FIG. 3

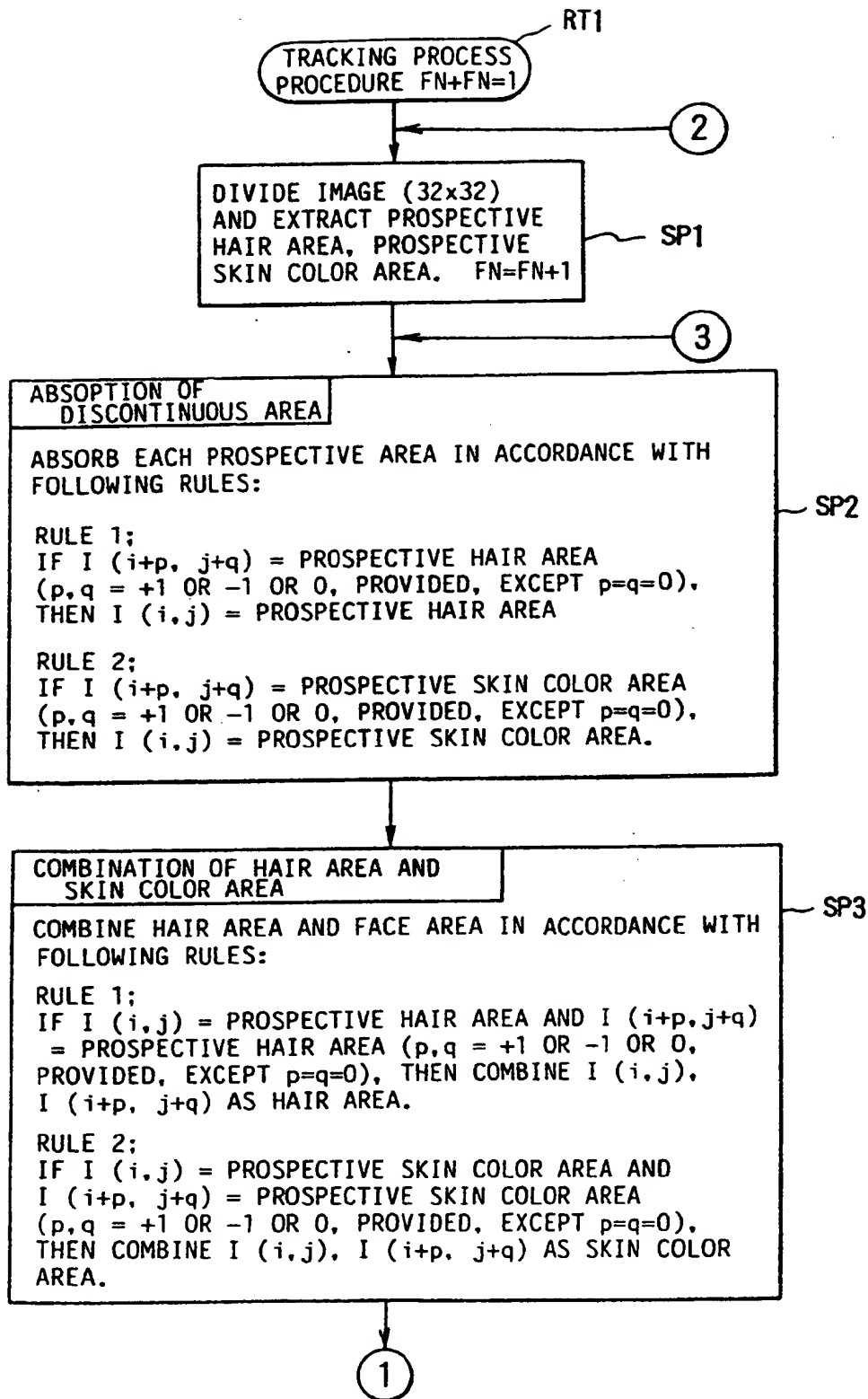


FIG. 4

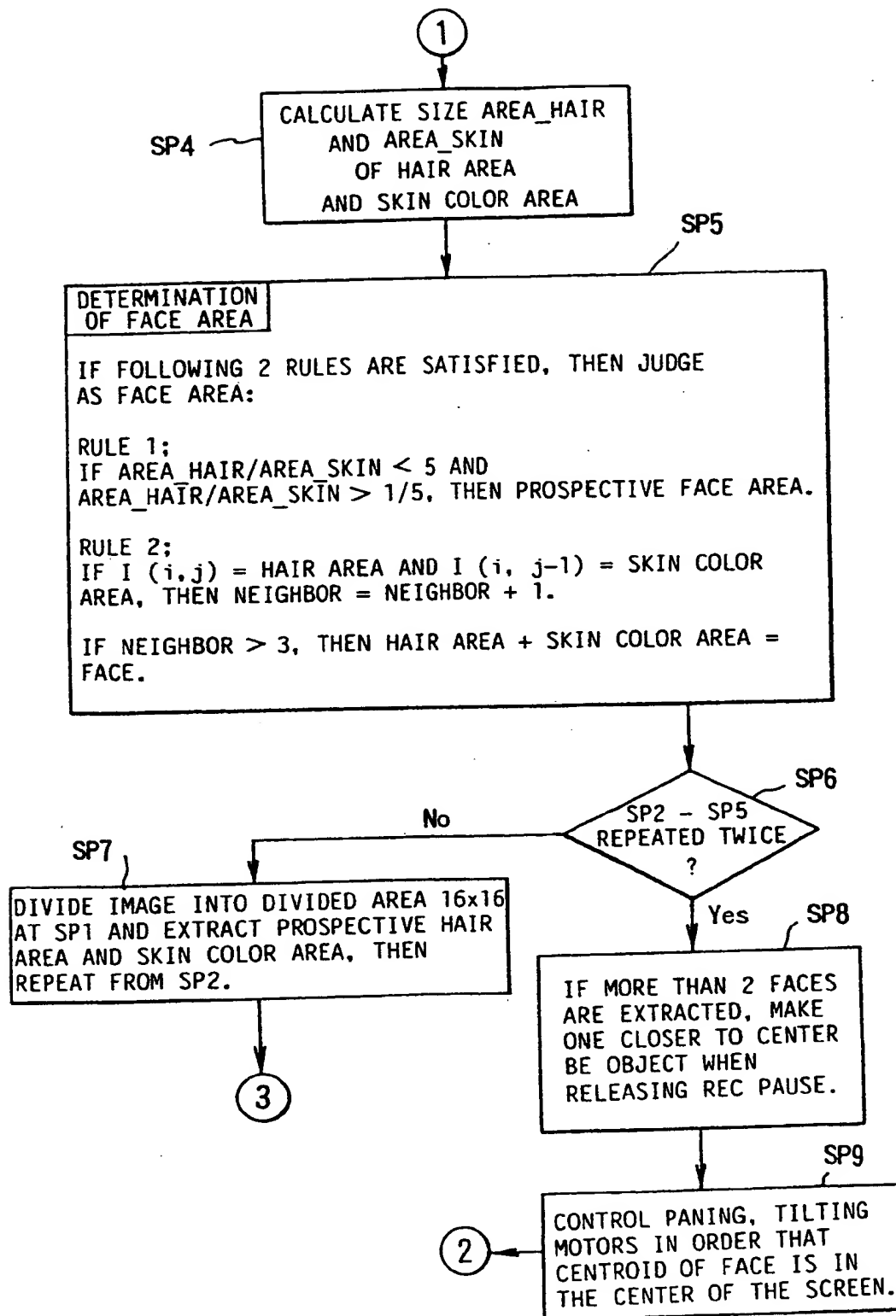


FIG. 5



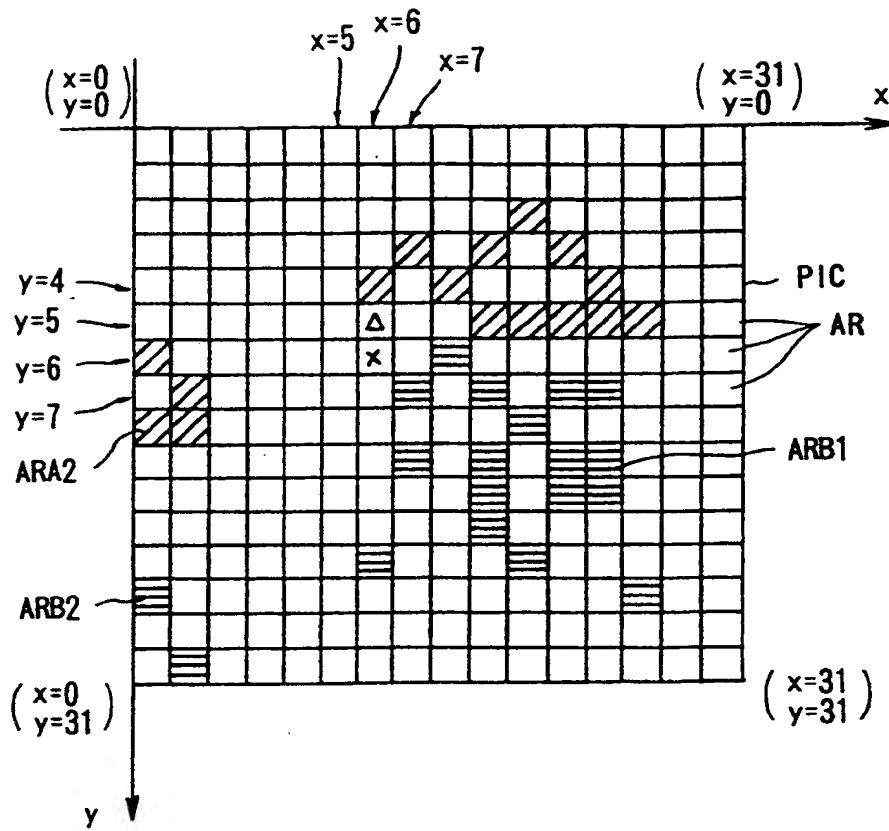


FIG. 6

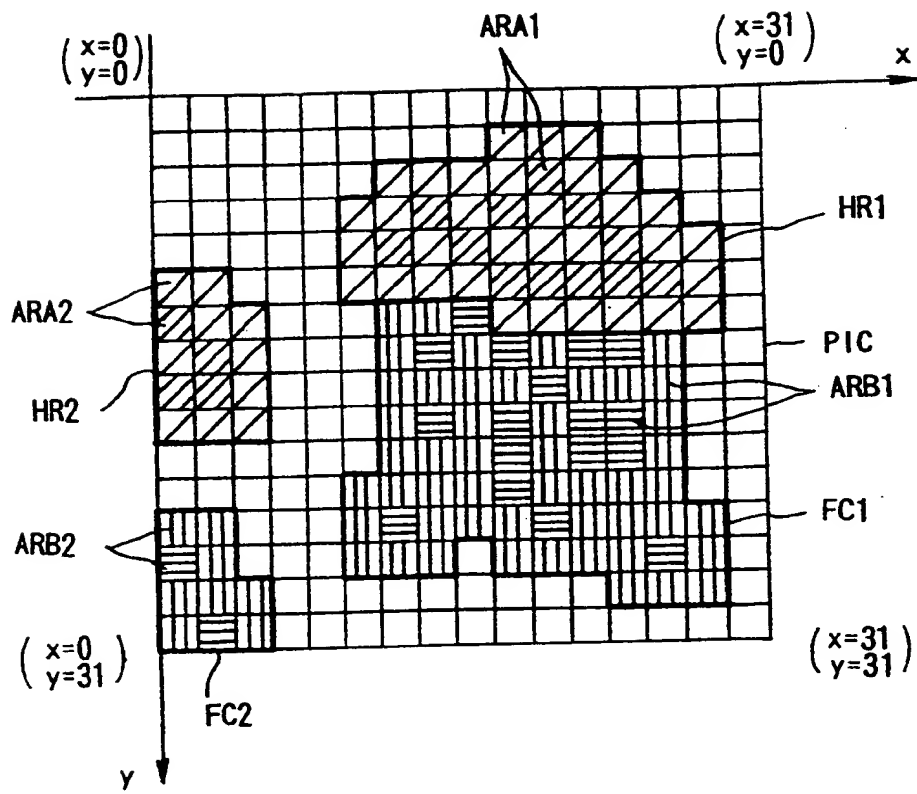


FIG. 7

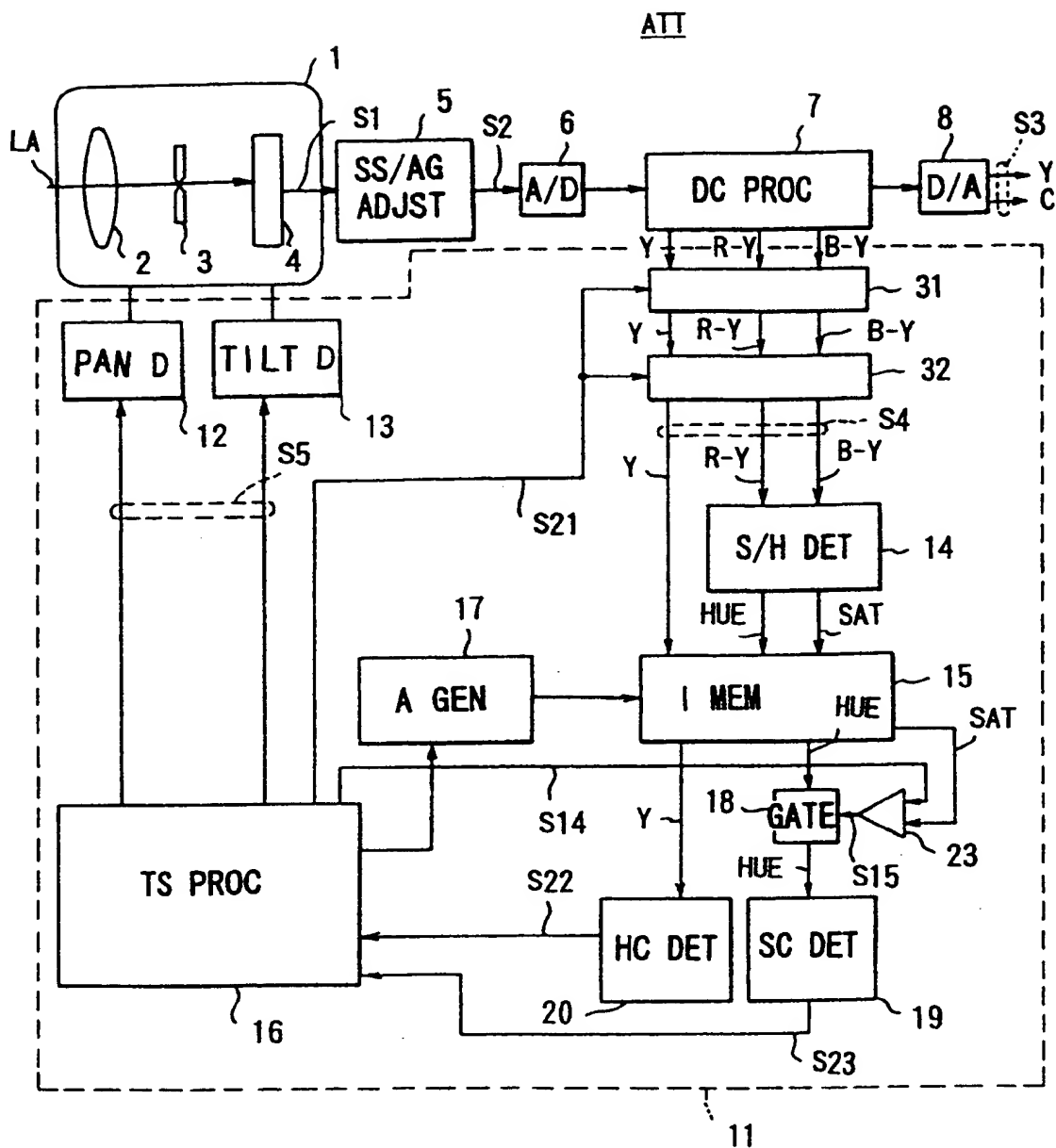


FIG. 8



①1 Publication number : **0 578 508 A3**

## EUROPEAN PATENT APPLICATION

②1 Application number : **93305419.9**

⑤1 Int. Cl.<sup>5</sup> : **H04N 5/232, G01S 3/786**

②2 Date of filing : **09.07.93**

③0 Priority : **10.07.92 JP 207107/92**

④3 Date of publication of application :  
**12.01.94 Bulletin 94/02**

⑧4 Designated Contracting States :  
**DE FR GB**

⑧8 Date of deferred publication of search report :  
**04.01.95 Bulletin 95/01**

⑦1 Applicant : **SONY CORPORATION**  
**6-7-35 Kitashinagawa**  
**Shinagawa-ku**  
**Tokyo 141 (JP)**

⑦2 Inventor : **Tomitaka, Tadafusa, c/o Legal & Intell. Prop. Dept**  
**Sony Corporation,**  
**6-7-35 Kitashinagawa**  
**Shinagawa-ku, Tokyo 141 (JP)**

⑦4 Representative : **Williams, Janice et al**  
**D. Young & Co.,**  
**21 New Fetter Lane**  
**London EC4A 1DA (GB)**

⑤4 **Video camera with colour-based target tracking system.**

⑤7 A video camera system comprising a teaching apparatus in which the field of view of a video camera can track the center of an object such a human face.

A display screen is divided into a lightness divided area in which a brightness signal has the prescribed luminance, and a skin color divided area in which a hue signal is skin color. Depending upon the extraction result, the position on the screen of the object is detected. A measuring frame which is not be affected by intensity and direction of the light source can then be set utilizing the hue signal.

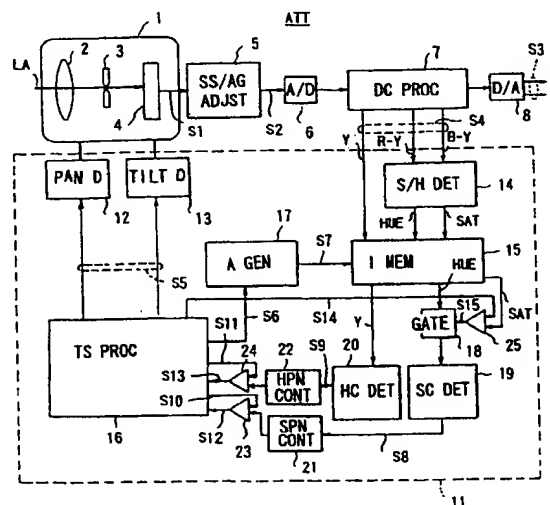


FIG. 1



European Patent  
Office

# EUROPEAN SEARCH REPORT

Application Number  
EP 93 30 5419

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. CL.5)
A	EP-A-0 439 334 (FUJITSU LIMITED) * abstract; figure 3 *	1-9	H04N5/232 G01S3/786
P, A	PATENT ABSTRACTS OF JAPAN vol. 17, no. 219 (E-1358) 30 April 1993 & JP-A-04 354 490 (MITSUBISHI HEAVY IND LTD) 8 December 1992 * abstract *	1-9	
A	US-A-5 093 869 (ALVES ET AL.) * column 1, line 32 - line 60 *	1-9	
A	PATENT ABSTRACTS OF JAPAN vol. 10, no. 127 (P-455) 13 May 1986 & JP-A-60 253 887 (CANON KK) 14 December 1985 * abstract *	1-9	
The present search report has been drawn up for all claims			<b>TECHNICAL FIELDS SEARCHED (Int. CL.5)</b>  H04N G01S G06K
Place of search <b>THE HAGUE</b>		Date of completion of the search <b>8 November 1994</b>	Examiner <b>Montanari, M</b>
<b>CATEGORY OF CITED DOCUMENTS</b> X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document			

EPO FORM 1500 (01.92) (P04.001)